Computational approach to quantum encoder design for purity optimization

Naoki Yamamoto[*]
Department of Engineering, Australian National University, ACT 0200, Australia

Maryam Fazel[†]
Control and Dynamical Systems, California Institute of Technology, Pasadena, California 91125, USA

(Dated: February 1, 2008)

In this paper, we address the problem of designing a quantum encoder that maximizes the minimum output purity of a given decohering channel, where the minimum is taken over all possible pure inputs. This problem is cast as a max-min optimization problem with a rank constraint on an appropriately defined matrix variable. The problem is computationally very hard because it is non-convex with respect to both the objective function (output purity) and the rank constraint. Despite this difficulty, we provide a tractable computational algorithm that produces the exact optimal solution for codespace of dimension two. Moreover, this algorithm is easily extended to cover the general class of codespaces, in which case the solution is suboptimal in the sense that the suboptimized output purity serves as a lower bound of the exact optimal purity. The algorithm consists of a sequence of semidefinite programmings and can be performed easily. Two typical quantum error channels are investigated to illustrate the effectiveness of our method.

PACS numbers: 03.67.Pp, 02.60.Pn

I. INTRODUCTION

The efficient transmission of quantum states over a noisy channel is a central subject in quantum information technologies [1]. The mathematical description of a quantum input-output relation is as follows. Let $\mathcal{H}$ and $\mathcal{K}$ be finite-dimensional Hilbert spaces of an input quantum state and the corresponding output, respectively. We denote by $\mathcal{L}(\mathcal{H},\mathcal{K})$ the set of linear operators from $\mathcal{H}$ to $\mathcal{K}$, and $\mathcal{S}(\mathcal{H})$ the set of quantum states on $\mathcal{H}$. The Markovian evolution of a quantum state $\rho \in \mathcal{S}(\mathcal{H})$ through a quantum channel $\mathcal{A}$ is typically modeled using the Kraus representation [2] as

$$\rho' = \mathcal{A}\rho = \sum_i A_i \rho A_i^\dagger, \tag{1}$$

where the Kraus operators $A_i \in \mathcal{L}(\mathcal{H},\mathcal{K})$ satisfy $\sum_i A_i^\dagger A_i = I_{\mathcal{H}}$ with $I_{\mathcal{H}}$ denoting the identity operator on $\mathcal{H}$. The *purity* of a state ρ is defined as $p[\rho] := \mathrm{Tr}\,(\rho^2)$, which is equal to one if and only if ρ is pure. Due to the decoherence caused by $\mathcal{A}$, a pure input state $\rho = |\phi\rangle\langle\phi|$ may be transmitted to a non-pure output $\rho' = \mathcal{A}(|\phi\rangle\langle\phi|)$ with $p[\rho'] < 1$. It is considered that $p[\rho']$ quantifies an intrinsic measure of the amount of decoherence induced by the error channel $\mathcal{A}$. In particular, this paper focuses on the *optimal purity*:

$$P(\mathcal{A}) := \max_{\mathcal{C}\subset\mathcal{H}} \min_{|\phi_c\rangle\in\mathcal{C}} \mathrm{Tr}\left[\mathcal{A}(|\phi_c\rangle\langle\phi_c|)^2\right], \tag{2}$$

where the minimization with respect to the state $|\phi_c\rangle$ takes into account the worst-case scenario of information processing. The maximization with respect to the *codespace* $\mathcal{C}\subset\mathcal{H}$ is motivated by the fact that we often have an opportunity to decrease the effect of decoherence by encoding our information into a higher-dimensional space; this is suggested by the theories of quantum error correction (QEC) [1, 3, 4, 5] and decoherence-free subspace (DFS) [6, 7, 8]. For example, embedding an input state $|\phi\rangle = \phi_1|0\rangle + \phi_2|1\rangle \in \mathbb{C}^2$ into a codespace spanned by $|00\rangle$ and $|11\rangle$ through the encoding process

$$\mathbb{C}^2 \ni |\phi\rangle \to |\phi_c\rangle = \phi_1|00\rangle + \phi_2|11\rangle \in \mathcal{C}\subset\mathcal{H} = \mathbb{C}^4 \tag{3}$$

appears to improve the output purity. Clearly, the most desirable situation is the existence of a DFS, i.e., a codespace that satisfies $P(\mathcal{A}) = 1$; but unfortunately this is a rare case. In this sense, the optimal codespace $\mathcal{C}$ is regarded as the best possible approximation of a DFS.

However, the max-min problem (2) is very hard to solve because it is non-convex with respect to both $\mathcal{C}$ and $|\phi_c\rangle$. To understand the structure of $P(\mathcal{A})$, in [9] Zanardi and Lidar considered channel purity for a fixed codespace $\mathcal{C}$ as

$$P(\mathcal{A},\mathcal{C}) := \min_{|\phi_c\rangle\in\mathcal{C}} \mathrm{Tr}\left[\mathcal{A}(|\phi_c\rangle\langle\phi_c|)^2\right], \tag{4}$$

and derived the alternative expression

$$P(\mathcal{A},\mathcal{C}) = \min_{|\phi_c\rangle\in\mathcal{C}} \langle\phi_c|\otimes\langle\phi_c|\Omega(\mathcal{A})|\phi_c\rangle\otimes|\phi_c\rangle,$$

where the Hermitian operator $\Omega(\mathcal{A})$ is defined by

$$\Omega(\mathcal{A}) := \sum_{ij}(A_j^\dagger A_i)\otimes(A_i^\dagger A_j) \in \mathcal{L}(\mathcal{H}^{\otimes 2},\mathcal{H}^{\otimes 2}). \tag{5}$$

This expression was used to derive a bound on $P(\mathcal{A},\mathcal{C})$ in terms of $\Omega(\mathcal{A})$ and $\mathcal{C}$, using techniques to calculate the expectation value of the "Hamiltonian" $\Omega(\mathcal{A})$. In the special case where eigenvectors of $\Omega(\mathcal{A})$ are product states in

[*]Electronic address: naoki.yamamoto@anu.edu.au
[†]Electronic address: maryam@cds.caltech.edu

a symmetric subspace of $\mathcal{H}^{\otimes 2}$, analytical expressions for $P(\mathcal{A}, \mathcal{C})$ were obtained. However, in general the max-min problem (2) does not have an analytical solution, leading us to take a computational approach.

From a computational point of view, owing to the rapid progress of computers, there have been many recent advances with a great potential for solving important problems in quantum theory. Convex optimization, and in particular semidefinite programming (SDP) [10, 11], have proven useful for quantum optimization problems such as a test for distinguishing an entangled from a separable quantum state [12, 13, 14, 15, 16] and a design of optimal measurement in linear quantum systems [17]. In addition, in [18, 19, 20] some quantum error-correction problems were solved using SDP, taking advantage of the well-known convexity of a set of quantum channels known as the *Jamiolkowski isomorphism* [21].

In this paper, we first use the same convexity property to set up a non-convex optimization problem that captures our goal and all the constraints. Then, we provide an algorithm that computes an *exact* local optimal solution of the hard non-convex problem (2) for the codespace of $\dim \mathcal{C} = 2$. This implies that the exact global optimal solution of (2) can be obtained by appropriately choosing an initial condition of the algorithm. The algorithm is represented by an iterative SDP and is thus computationally tractable. The derivation of the SDP consists of two stages. The first one transforms the constraints to equivalent Linear Matrix Inequality (LMI) constraints. The key idea used to obtain the LMI in this stage is the *Sum-of-Squares* characterization of a polynomial constraint [22, 23, 24]. In the second stage, a non-convex rank constraint of the matrix variable is tackled via the *log-det* (logarithm of determinant) heuristic [25, 26, 27]. Furthermore, we will show an extended version of the above SDP algorithm that computes a *lower bound* of the optimal purity $P(\mathcal{A})$ for the general class of $\mathcal{C}$.

This paper is organized as follows. Section II reviews the Jamiolkowski isomorphism, which is used to formulate the optimization problem in Section III. The SDP algorithm is presented in Section IV. The general case that leads to a suboptimal solution is discussed in Section V. In Section VI, we examine two typical quantum error channels, the *bit-flip channel* and the *amplitude damping channel*, and demonstrate the effectiveness of our method. Section VII concludes the paper.

Notation: A Hermitian matrix $X = X^\dagger \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^n)$ is *positive semidefinite* if $\langle a|X|a\rangle \geq 0,\ \forall |a\rangle \in \mathbb{C}^n$; the inequality $X \geq 0$ represents the positive semidefiniteness of X. We use I_n to denote the $n \times n$ identity matrix, which is the same as $I_\mathcal{H}$ when $\dim \mathcal{H} = n$. For a matrix $X = (x_{ij})$, the symbols X^T and X^* represent the matrix transpose and the elementwise complex conjugate of X, i.e., $X^\mathsf{T} = (x_{ji})$ and $X^* = (x_{ij}^*) = (X^\dagger)^\mathsf{T}$, respectively; these rules are applied to any rectangular matrix including column and row vectors. $\Re(X)$ and $\Im(X)$ denote the real and imaginary part of X, respectively, i.e., $(\Re(X))_{ij} = (x_{ij} + x_{ij}^*)/2$ and $(\Im(X))_{ij} = (x_{ij} - x_{ij}^*)/2\mathrm{i}$.

II. THE JAMIOLKOWSKI ISOMORPHISM

The main purpose of this section is to review the following important fact known as the Jamiolkowski isomorphism [21]; the set of all finite-dimensional quantum channels has a one-to-one correspondence with a convex set of positive semidefinite matrices acting on $\mathcal{K} \otimes \mathcal{H}$. This fact can be seen in various ways [18, 28, 29, 30]. Here we follow the notations in [18, 30] and obtain two matrix representations of a quantum channel, which we later use to set up the optimization problem. At the end of this section, we present a characterization of quantum channels that preserve pure states.

We consider a general trace-preserving quantum channel that maps an input $\rho \in \mathcal{S}(\mathcal{H}) = \mathcal{S}(\mathbb{C}^n)$ to the output

$$\rho' = \sum_i X_i \rho X_i^\dagger \in \mathcal{S}(\mathcal{K}) = \mathcal{S}(\mathbb{C}^m). \tag{6}$$

Let $\{|i\rangle\}_{i=1,\cdots,n}$ and $\{|\bar{i}\rangle\}_{i=1,\cdots,m}$ be orthonormal bases in $\mathcal{H}$ and $\mathcal{K}$, respectively. Then, any vectors in $\mathcal{H}^{\otimes 2}$ and $\mathcal{K}^{\otimes 2}$ are expressed as $|\Phi\rangle\rangle = \sum_{i,j=1}^n \phi_{ij} |i\rangle \otimes |j\rangle$ and $|\Phi'\rangle\rangle = \sum_{i,j=1}^m \phi'_{ij} |\bar{i}\rangle \otimes |\bar{j}\rangle$, respectively. We sometimes use $|i\rangle|j\rangle$ as a short-hand for $|i\rangle \otimes |j\rangle$. Let us now define the following two specific vectors:

$$|I_\mathcal{H}\rangle\rangle := \sum_{i=1}^n |i\rangle \otimes |i\rangle^* \in \mathcal{H}^{\otimes 2}, \tag{7}$$

$$|I_\mathcal{K}\rangle\rangle := \sum_{i=1}^m |\bar{i}\rangle \otimes |\bar{i}\rangle^* \in \mathcal{K}^{\otimes 2}. \tag{8}$$

These vectors have the property of being independent of the selection of orthonormal basis; for any two orthonormal bases $\{|a_i\rangle\}$ and $\{|b_i\rangle\}$ in $\mathcal{H}$, we have

$$|I_\mathcal{H}\rangle\rangle = \sum_{i=1}^n |a_i\rangle \otimes |a_i\rangle^* = \sum_{i=1}^n |b_i\rangle \otimes |b_i\rangle^*. \tag{9}$$

Note that the invariant property (9) is not satisfied if $|I_\mathcal{H}\rangle\rangle$ is defined without the complex conjugation. The vectors (7) and (8) are related by

$$(X \otimes I_\mathcal{H})|I_\mathcal{H}\rangle\rangle = (I_\mathcal{K} \otimes X^\mathsf{T})|I_\mathcal{K}\rangle\rangle, \quad \forall X \in \mathcal{L}(\mathcal{H}, \mathcal{K}). \tag{10}$$

Further, the following equation holds:

$$\langle\langle I_\mathcal{H}|(X \otimes I_\mathcal{H})|I_\mathcal{H}\rangle\rangle = \mathrm{Tr}\, X, \quad \forall X \in \mathcal{L}(\mathcal{H}, \mathcal{H}). \tag{11}$$

We now define a positive semidefinite matrix $\boldsymbol{X}_1$ associated with the Kraus operators $X_i \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ as

$$\boldsymbol{X}_1 := \sum_i (X_i \otimes I_\mathcal{H})|I_\mathcal{H}\rangle\rangle\langle\langle I_\mathcal{H}|(X_i \otimes I_\mathcal{H})^\dagger$$
$$\in \mathcal{L}(\mathcal{K} \otimes \mathcal{H}, \mathcal{K} \otimes \mathcal{H}).$$

Then, the trace-preserving condition $\sum_i X_i^\dagger X_i = I_\mathcal{H}$ corresponds to $\mathrm{Tr}\,_\mathcal{K} \boldsymbol{X}_1 = I_\mathcal{H}$, and the quantum channel (6) is expressed in terms of $\boldsymbol{X}_1$ as

$$\rho' = \mathrm{Tr}\,_\mathcal{H}\Big[(I_\mathcal{K} \otimes \rho^\mathsf{T})\boldsymbol{X}_1\Big]. \tag{12}$$

Conversely, it is known that any positive semidefinite matrix $\boldsymbol{X}_1 \in \mathcal{L}(\mathcal{K}\otimes\mathcal{H},\mathcal{K}\otimes\mathcal{H})$ corresponds to a quantum channel with input-output relation given in Eq. (12). That is, there exists a one-to-one correspondence between a quantum channel from $\mathcal{H}$ to $\mathcal{K}$ and a positive semidefinite matrix on $\mathcal{K}\otimes\mathcal{H}$.

We next introduce another matrix representation of the quantum channel, which will be denoted by $\boldsymbol{X}_2$. To this end, we define a vector associated with a quantum state $\rho\in\mathcal{S}(\mathcal{H})$ as

$$|\rho\rangle\!\rangle := (\rho\otimes I_{\mathcal{H}})|I_{\mathcal{H}}\rangle\!\rangle \in \mathcal{H}^{\otimes 2}. \tag{13}$$

The vector $|\rho\rangle\!\rangle$ is obviously in one-to-one correspondence with ρ. In particular, from Eq. (9), the vector representation of a pure state $\rho=|a\rangle\langle a|$ is given by

$$|\rho\rangle\!\rangle = (|a\rangle\langle a|\otimes I_{\mathcal{H}})\sum_i |i\rangle\otimes|i\rangle^* = |a\rangle\otimes|a\rangle^*. \tag{14}$$

In addition, the purity $p[\rho]=\mathrm{Tr}\,(\rho^2)$ is simply the squared Euclidean norm of $|\rho\rangle\!\rangle$:

$$p[\rho] = \mathrm{Tr}\,(\rho^2) = \langle\!\langle\rho|\rho\rangle\!\rangle, \tag{15}$$

due to Eq. (11). Thus, a quantum state $|\rho\rangle\!\rangle$ is pure if and only if $\langle\!\langle\rho|\rho\rangle\!\rangle=1$. Let us now define $\boldsymbol{X}_2$. Multiplying $|I_{\mathcal{K}}\rangle\!\rangle$ on both sides of Eq. (6), we have $(\rho'\otimes I_{\mathcal{K}})|I_{\mathcal{K}}\rangle\!\rangle = \sum_i (X_i\rho X_i^\dagger\otimes I_{\mathcal{K}})|I_{\mathcal{K}}\rangle\!\rangle$, which is rewritten by

$$\begin{aligned}(\rho'\otimes I_{\mathcal{K}})|I_{\mathcal{K}}\rangle\!\rangle &= \sum_i (X_i\otimes I_{\mathcal{K}})(\rho\otimes I_{\mathcal{K}})(I_{\mathcal{H}}\otimes X_i^*)|I_{\mathcal{H}}\rangle\!\rangle\\ &= \sum_i (X_i\otimes X_i^*)(\rho\otimes I_{\mathcal{H}})|I_{\mathcal{H}}\rangle\!\rangle,\end{aligned}$$

because of the property (10). Hence, defining the matrix

$$\boldsymbol{X}_2 := \sum_i X_i\otimes X_i^* \in \mathcal{L}(\mathcal{H}^{\otimes 2},\mathcal{K}^{\otimes 2}),$$

the quantum channel (6) is represented by

$$\mathcal{H}^{\otimes 2}\ni|\rho\rangle\!\rangle \to |\rho'\rangle\!\rangle = \boldsymbol{X}_2|\rho\rangle\!\rangle \in \mathcal{K}^{\otimes 2}.$$

The trace-preserving condition is then given by

$$\begin{aligned}\langle\!\langle I_{\mathcal{K}}|\boldsymbol{X}_2 &= \sum_i \langle\!\langle I_{\mathcal{K}}|(X_i\otimes I_{\mathcal{K}})(I_{\mathcal{H}}\otimes X_i^*)\\ &= \sum_i \langle\!\langle I_{\mathcal{H}}|(I_{\mathcal{H}}\otimes X_i^{\mathsf{T}})(I_{\mathcal{H}}\otimes X_i^*) = \langle\!\langle I_{\mathcal{H}}|.\end{aligned}$$

The matrix $\boldsymbol{X}_2$ is related to $\boldsymbol{X}_1$ through the following rearrangement rule of the matrix elements:

$$\langle\bar{i}|\langle\bar{j}|^*\boldsymbol{X}_2|k\rangle|\ell\rangle^* = \langle\bar{i}|\langle k|^*\boldsymbol{X}_1|\bar{j}\rangle|\ell\rangle^*.$$

This relation is independent of the selection of $\{|i\rangle\}$ and $\{|\bar{i}\rangle\}$ due to Eq. (9). As the rearrangement map is obviously linear and homeomorphic, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have a one-to-one correspondence with each other. We denote this relation by $\boldsymbol{X}_1=\Phi(\boldsymbol{X}_2)$. The above discussion is summarized as follows.

Lemma 1. Any finite-dimensional quantum channel from $\mathcal{H}$ to $\mathcal{K}$ is represented by $\mathcal{H}^{\otimes 2}\ni|\rho\rangle\!\rangle\to|\rho'\rangle\!\rangle = \boldsymbol{X}|\rho\rangle\!\rangle\in\mathcal{K}^{\otimes 2}$, where $\boldsymbol{X}$ is in the set

$$\mathcal{X}(\mathcal{H},\mathcal{K}) = \Big\{\ \boldsymbol{X}\in\mathcal{L}(\mathcal{H}^{\otimes 2},\mathcal{K}^{\otimes 2})\ \Big|\ \Phi(\boldsymbol{X})\geq 0,\ \langle\!\langle I_{\mathcal{K}}|\boldsymbol{X} = \langle\!\langle I_{\mathcal{H}}|\ \Big\}.$$

The linear transformation $\Phi(\boldsymbol{X})$ is defined with respect to orthonormal bases $\{|i\rangle\}\in\mathcal{H}$ and $\{|\bar{i}\rangle\}\in\mathcal{K}$ as

$$\langle\bar{i}|\langle\bar{j}|^*\boldsymbol{X}|k\rangle|\ell\rangle^* = \langle\bar{i}|\langle k|^*\Phi(\boldsymbol{X})|\bar{j}\rangle|\ell\rangle^*.$$

Clearly, $\mathcal{X}(\mathcal{H},\mathcal{K})$ is a convex set with dimension $m^2n^2-n^2$. It should be noted that a cascade connection of two quantum channels $\boldsymbol{X}\in\mathcal{X}(\mathcal{H},\mathcal{K})$ and $\boldsymbol{Y}\in\mathcal{X}(\mathcal{K},\mathcal{V})$ is simply represented by the multiplication of those matrices: $\boldsymbol{Y}\boldsymbol{X}\in\mathcal{X}(\mathcal{H},\mathcal{V})$.

Finally, we provide a characterization of quantum channels that preserve pure states, i.e., $p[\rho]=p[\rho']=1$, as follows.

Lemma 2. For a quantum channel $\boldsymbol{X}\in\mathcal{X}(\mathcal{H},\mathcal{K})$, the following three conditions are equivalent.

(i) $\boldsymbol{X}|a\rangle\!\rangle$ is pure for any pure state $|a\rangle\!\rangle = |a\rangle\otimes|a\rangle^*$.

(ii) $\boldsymbol{X}^\dagger\boldsymbol{X} = I_{\mathcal{H}^{\otimes 2}} = I_{n^2}$

(iii) $\mathrm{rank}\Phi(\boldsymbol{X})=1$

Proof. (i) ⇔ (ii). Condition (ii) immediately implies that $|a'\rangle\!\rangle = \boldsymbol{X}|a\rangle\!\rangle$ is pure, since $p[a'] = \langle\!\langle a'|a'\rangle\!\rangle = \langle\!\langle a|\boldsymbol{X}^\dagger\boldsymbol{X}|a\rangle\!\rangle = \langle\!\langle a|a\rangle\!\rangle = 1$. Conversely, as $\boldsymbol{X}$ can be represented by $\boldsymbol{X}=\sum_{i=1}^M X_i\otimes X_i^*$, the quantum state $|a'\rangle\!\rangle = \boldsymbol{X}|a\rangle\!\rangle$ always satisfies the following relation:

$$\begin{aligned}\langle\!\langle a'|a'\rangle\!\rangle &= \langle\!\langle a|\boldsymbol{X}^\dagger\boldsymbol{X}|a\rangle\!\rangle\\ &= \langle a|\langle a|^*\sum_{i,j}(X_i^\dagger\otimes X_i^{\mathsf{T}})(X_j\otimes X_j^*)|a\rangle|a\rangle^*\\ &= \sum_{i,j}|\langle a|X_i^\dagger X_j|a\rangle|^2\\ &\leq \sum_{i,j}\langle a|X_i^\dagger X_i|a\rangle\langle a|X_j^\dagger X_j|a\rangle = 1.\end{aligned} \tag{16}$$

Therefore, the condition $\langle\!\langle a'|a'\rangle\!\rangle=1$ imposes the equality relation in Eq. (16). Then, $X_i|a\rangle$ is parallel to $X_j|a\rangle$ for all (i,j) and $|a\rangle$, indicating that X_i is independent of i. Thus, $\boldsymbol{X}$ takes the form $\boldsymbol{X}=X\otimes X^*$, where X is defined by $X:=\sqrt{M}X_i$. Consequently, using the trace-preserving condition $X^\dagger X = I_{\mathcal{H}}$, we arrive at $\boldsymbol{X}^\dagger\boldsymbol{X} = I_{n^2}$.

(ii) ⇔ (iii). First, we assume (iii). Then, $\Phi(\boldsymbol{X})$ is written as $\Phi(\boldsymbol{X}) = |x\rangle\!\rangle\langle\!\langle x|$ using a vector $|x\rangle\!\rangle\in\mathcal{K}\otimes\mathcal{H}$. Furthermore, as $|x\rangle\!\rangle$ can be represented by $|x\rangle\!\rangle = (X\otimes I_{\mathcal{H}})|I_{\mathcal{H}}\rangle\!\rangle$ with a matrix $X\in\mathcal{L}(\mathcal{H}\otimes\mathcal{K})$, we have $\Phi(\boldsymbol{X}) = (X\otimes I_{\mathcal{H}})|I_{\mathcal{H}}\rangle\!\rangle\langle\!\langle I_{\mathcal{H}}|(X\otimes I_{\mathcal{H}})^\dagger$, and thus $\boldsymbol{X}=X\otimes X^*$ from the definition of Φ. This directly yields $\boldsymbol{X}^\dagger\boldsymbol{X}=I_{n^2}$ due to $X^\dagger X = I_{\mathcal{H}}$. We next turn to the proof of (ii) ⇒ (iii). Multiplying a pure state $|a\rangle\rangle = |a\rangle \otimes |a\rangle^* \in \mathcal{H}^{\otimes 2}$ on both sides of $I_{n^2} = \boldsymbol{X}^\dagger \boldsymbol{X}$ where $\boldsymbol{X} = \sum_i X_i \otimes X_i^*$, we obtain

$$\begin{aligned} 1 = \langle\langle a|a\rangle\rangle = \langle\langle a|\boldsymbol{X}^\dagger \boldsymbol{X}|a\rangle\rangle &= \sum_{i,j} |\langle a|X_i^\dagger X_j|a\rangle|^2 \\ &\le \sum_{i,j} \langle a|X_i^\dagger X_i|a\rangle \langle a|X_j^\dagger X_j|a\rangle = 1. \end{aligned}$$

Hence, from the same reason as in the proof of (i) ⇒ (ii), $\boldsymbol{X}$ must be of the form $\boldsymbol{X} = X \otimes X^*$. This implies $\Phi(\boldsymbol{X}) = (X \otimes I_{\mathcal{H}})|I_{\mathcal{H}}\rangle\rangle\langle\langle I_{\mathcal{H}}|(X \otimes I_{\mathcal{H}})^\dagger$ and thus $\mathrm{rank}\Phi(\boldsymbol{X}) = 1$. ■

Corollary 3. Suppose $\boldsymbol{X} \in \mathcal{X}(\mathcal{H}, \mathcal{K})$ satisfies $\mathrm{rank}\Phi(\boldsymbol{X}) = 1$. Then, the nonzero eigenvalue of $\Phi(\boldsymbol{X})$ is given by $n = \dim \mathcal{H}$.

Proof. From the proof of Lemma 2, we have

$$\mathrm{Tr}\, \Phi(\boldsymbol{X}) = \langle\langle I_{\mathcal{H}}|(X^\dagger X \otimes I_{\mathcal{H}})|I_{\mathcal{H}}\rangle\rangle = \langle\langle I_{\mathcal{H}}|I_{\mathcal{H}}\rangle\rangle = n. \ ■$$

According to Lemma 2, the totality of quantum channels that transform pure states in $\mathcal{H}$ to pure in $\mathcal{K}$ is completely characterized by the following non-convex set:

$$\begin{aligned} \mathcal{X}_1(\mathcal{H}, \mathcal{K}) = \Big\{ \boldsymbol{X} \in \mathcal{L}(\mathcal{H}^{\otimes 2}, \mathcal{K}^{\otimes 2}) \ \Big| \ &\mathrm{rank}\Phi(\boldsymbol{X}) = 1, \\ &\Phi(\boldsymbol{X}) \ge 0, \ \langle\langle I_{\mathcal{K}}|\boldsymbol{X} = \langle\langle I_{\mathcal{H}}| \ \Big\}. \end{aligned}$$

III. OPTIMAL ENCODER DESIGN AS A MATRIX OPTIMIZATION PROBLEM

This section is devoted to rewrite the problem (2) as an encoder-optimization problem, which is further described as a matrix optimization problem using the notations introduced in Section II.

First, let us fix the dimension of the codespace $\mathcal{C}$ to $\dim \mathcal{C} = r$ and represent an element of $\mathcal{C}$ by $|\phi_c\rangle = E|\phi\rangle$ with the input pure state $|\phi\rangle \in \mathbb{C}^r$ which contains all information of the sender. Here, E is the Kraus operator corresponding to the following encoding channel:

$$\mathcal{E} : \ \mathbb{C}^r \ni |\phi\rangle \to |\phi_c\rangle = E|\phi\rangle \in \mathcal{C} \subset \mathcal{H}. \qquad (17)$$

We set $\mathcal{H} = \mathbb{C}^n$; then, E is an $n \times r$ complex matrix satisfying $E^\dagger E = I_r$. In the example (3), $|\phi\rangle$ is a qubit and $E = |00\rangle\langle 0| + |11\rangle\langle 1|$, i.e., $r = 2$ and $n = 4$. In terms of the above notations, the codespace-optimization problem (2) is written as

$$\begin{aligned} P(\mathcal{A}) &= \max_{\mathcal{E}} \min_{|\phi\rangle \in \mathbb{C}^r} P(\mathcal{A}, \mathcal{E}, |\phi\rangle), \\ P(\mathcal{A}, \mathcal{E}, |\phi\rangle) &= \mathrm{Tr}\left[\mathcal{A}\mathcal{E}(|\phi\rangle\langle\phi|)^2\right] \\ &= \mathrm{Tr}\left[\mathcal{A}(E|\phi\rangle\langle\phi|E^\dagger)^2\right]. \end{aligned} \qquad (18)$$

Next, let us represent the problem using the matrix variable introduced in Section II. Since the encoding channel $\mathcal{E}$ obviously preserves pure states, its matrix representation $\boldsymbol{E}$ is an element of $\mathcal{X}_1(\mathbb{C}^r, \mathbb{C}^n)$. Also, from Eq. (14), the input $|\phi\rangle$ takes the form $|\phi\rangle\rangle = |\phi\rangle \otimes |\phi\rangle^*$ in the extended space $(\mathbb{C}^r)^{\otimes 2}$. Hence, the output state of the encoder-error process is given by $|\rho'\rangle\rangle = \boldsymbol{A}\boldsymbol{E}|\phi\rangle|\phi\rangle^*$, where $\boldsymbol{A} \in \mathcal{X}(\mathbb{C}^n, \mathbb{C}^n)$ is the matrix representation of the error channel $\mathcal{A}$. Then, due to Eq. (15), the output purity is

$$P(\mathcal{A}, \mathcal{E}, |\phi\rangle) = \langle\langle \rho'|\rho'\rangle\rangle = \langle\phi|\langle\phi|^* \boldsymbol{E}^\dagger \boldsymbol{A}^\dagger \boldsymbol{A}\boldsymbol{E}|\phi\rangle|\phi\rangle^*.$$

Consequently, the max-min problem (18) is written as

$$P(\mathcal{A}) = \max_{\mathbf{E} \in \mathcal{X}_1} \min_{|\phi\rangle \in \mathbb{C}^r} \langle\phi|\langle\phi|^* \boldsymbol{E}^\dagger \boldsymbol{A}^\dagger \boldsymbol{A}\boldsymbol{E}|\phi\rangle|\phi\rangle^*, \qquad (19)$$

which is identical to the following "error-minimization" problem:

$$\begin{aligned} \min_{\mathbf{E}, \epsilon} \quad & \epsilon, \\ \text{s.t.} \quad & \langle\phi|\langle\phi|^* \boldsymbol{E}^\dagger \boldsymbol{A}^\dagger \boldsymbol{A}\boldsymbol{E}|\phi\rangle|\phi\rangle^* \ge 1 - \epsilon, \ \forall|\phi\rangle \in \mathbb{C}^r, \\ & \mathbf{E} \in \mathcal{X}_1(\mathbb{C}^r, \mathbb{C}^n), \\ & 0 \le \epsilon \le 1. \end{aligned} \qquad (20)$$

Note that the optimal purity is related to the minimum error, ϵ_{opt}, by

$$P(\mathcal{A}) = 1 - \epsilon_{\mathrm{opt}}.$$

IV. EXACT OPTIMAL SOLUTION TO THE PURITY-OPTIMIZATION PROBLEM

In this section, we provide a systematic and powerful computational algorithm that exactly solves the purity-optimization problem when $\dim \mathcal{C} = r = 2$. The proposed algorithm can easily be extended to cover the general class of codespaces of dimension $r \ge 3$, in which case the suboptimized output purity gives a lower bound of the optimal purity $P(\mathcal{A})$. This result will be discussed in Section V.

The procedure to derive the algorithm consists of two stages. In the first stage, it will be proved that the first constraint in the problem (20):

$$\langle\phi|\langle\phi|^* \boldsymbol{E}^\dagger \boldsymbol{A}^\dagger \boldsymbol{A}\boldsymbol{E}|\phi\rangle|\phi\rangle^* \ge 1 - \epsilon, \ \forall|\phi\rangle \in \mathbb{C}^2 \qquad (21)$$

can be equivalently transformed to an LMI condition with respect to $\boldsymbol{E}$, ϵ, and an additional variable. In the second stage, we will consider a tractable rank-minimization problem of the matrix variable that is closely related to the original error-minimization problem (20). It will be then shown that, under a certain condition, the optimal solution of the rank-minimization problem coincides with that of the problem (20).

A. The first stage: transformation of the constraint

To simplify the exposition, we here assume that the input state $|\phi\rangle$ is a real-valued qubit, i.e., $|\phi\rangle = [x_1, x_2]^\mathsf{T} \in$

$\mathbb{R}^2$, $(x_1^2 + x_2^2 = 1)$. The general qubit case $|\phi\rangle \in \mathbb{C}^2$ will be discussed in Section IV-C using essentially the same idea presented here.

Before considering the transformation of the constraint (21), let us further express it only in terms of real matrices. To this end, we define the following real matrix variable with the size $2n^2 \times 4$:

$$\tilde{\boldsymbol{E}} := \begin{bmatrix} \Re(\boldsymbol{E}) \\ \Im(\boldsymbol{E}) \end{bmatrix}. \tag{22}$$

Then, the output purity is expressed as

$$P(\mathcal{A}, \mathcal{E}, |\phi\rangle) = \langle\phi|\langle\phi|\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}|\phi\rangle|\phi\rangle,$$

where $\boldsymbol{P}$ is a real positive semidefinite matrix defined by

$$\boldsymbol{P} := \begin{bmatrix} \Re(\boldsymbol{A}^\dagger\boldsymbol{A}) & -\Im(\boldsymbol{A}^\dagger\boldsymbol{A}) \\ \Im(\boldsymbol{A}^\dagger\boldsymbol{A}) & \Re(\boldsymbol{A}^\dagger\boldsymbol{A}) \end{bmatrix}.$$

Furthermore, we introduce a matrix [32]

$$\boldsymbol{B} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{23}$$

and define a vector

$$|x\rangle\rangle_B := \boldsymbol{B}^{\mathsf{T}}|\phi\rangle|\phi\rangle = \boldsymbol{B}^{\mathsf{T}}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \otimes \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}.$$

Note that $|x\rangle\rangle_B$ is normalized: ${}_B\langle\langle x|x\rangle\rangle_B = 1$. As a result, from the relation $|\phi\rangle|\phi\rangle = \boldsymbol{B}|x\rangle\rangle_B$, the constraint (21) is expressed as

$$p(x) := {}_B\langle\langle x|\Big[\boldsymbol{B}^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{B} + (\epsilon - 1)I_3\Big]|x\rangle\rangle_B \geq 0, \quad \forall x_1, x_2 \in \mathbb{R}. \tag{24}$$

We are now in the position to describe the transformation. The constraint (24) indicates that $p(x)$ must be a real fourth-order nonnegative polynomial function with respect to the variables (x_1, x_2). This type of constraint, i.e., the nonnegativity of a polynomial function, frequently appears in a wide variety of engineering problems. In particular, the following Sum-of-Squares (SOS) characterization of non-negative polynomials, first studied by David Hilbert more than a century ago, is a fundamental question: When does a nonnegative polynomial $p(x)$ have an SOS decomposition $p(x) = \sum_i h_i^2(x)$ for some polynomials $h_i(x)$? One of the well-known answers to the above question leads us to conclude that the nonnegative polynomial $p(x)$ must have an SOS decomposition, thereby Eq. (24) is equivalently replaced by the following matrix inequality:

$$\begin{aligned} &\boldsymbol{B}^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{B} + (\epsilon - 1)I_3 + \tau\boldsymbol{S} \\ &+ \boldsymbol{T}_1^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{T}_2 + \boldsymbol{T}_2^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{T}_1 \\ &- \boldsymbol{T}_3^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{T}_4 - \boldsymbol{T}_4^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{T}_3 \geq 0, \end{aligned} \tag{25}$$

where $\tau \in \mathbb{R}$ is an additional optimization variable. The proof of Eq. (25) and the matrices $\boldsymbol{T}_1, \boldsymbol{T}_2, \boldsymbol{T}_3, \boldsymbol{T}_4$, and $\boldsymbol{S}$ are given in Appendix A. The inequality (25) is transformed to

$$\begin{aligned} &\tau\boldsymbol{S} + (\epsilon - 1)I_3 - \begin{bmatrix} \tilde{\boldsymbol{E}}\boldsymbol{B} \\ \tilde{\boldsymbol{E}}\boldsymbol{T}_1 \\ \tilde{\boldsymbol{E}}\boldsymbol{T}_2 \\ \tilde{\boldsymbol{E}}\boldsymbol{T}_3 \\ \tilde{\boldsymbol{E}}\boldsymbol{T}_4 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} kI_{2n^2} - \boldsymbol{P} & & & & \\ & kI_{2n^2} & -\boldsymbol{P} & & \\ & -\boldsymbol{P} & kI_{2n^2} & & \\ & & & kI_{2n^2} & \boldsymbol{P} \\ & & & \boldsymbol{P} & kI_{2n^2} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{E}}\boldsymbol{B} \\ \tilde{\boldsymbol{E}}\boldsymbol{T}_1 \\ \tilde{\boldsymbol{E}}\boldsymbol{T}_2 \\ \tilde{\boldsymbol{E}}\boldsymbol{T}_3 \\ \tilde{\boldsymbol{E}}\boldsymbol{T}_4 \end{bmatrix} \\ &+ k\Big[\boldsymbol{B}^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\tilde{\boldsymbol{E}}\boldsymbol{B} + \boldsymbol{T}_1^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\tilde{\boldsymbol{E}}\boldsymbol{T}_1 + \boldsymbol{T}_2^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\tilde{\boldsymbol{E}}\boldsymbol{T}_2 + \boldsymbol{T}_3^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\tilde{\boldsymbol{E}}\boldsymbol{T}_3 + \boldsymbol{T}_4^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\tilde{\boldsymbol{E}}\boldsymbol{T}_4\Big] \geq 0, \end{aligned} \tag{26}$$

where the blank spaces in the large matrix denote zero entries. The fixed scalar number $k > 0$ is selected such that

$$kI_{2n^2} - \boldsymbol{P} > 0 \tag{27}$$

is satisfied. This is equivalent to

$$\begin{bmatrix} kI_{2n^2} & \boldsymbol{P} \\ \boldsymbol{P} & kI_{2n^2} \end{bmatrix} > 0, \ \begin{bmatrix} kI_{2n^2} & -\boldsymbol{P} \\ -\boldsymbol{P} & kI_{2n^2} \end{bmatrix} > 0. \tag{28}$$

Then, due to the conditions (27) and (28), the large matrix in Eq. (26) is positive definite. Moreover, we now see from Lemma 2 that the non-convex rank condition $\mathrm{rank}\Phi(\boldsymbol{E}) = 1$ is equivalent to $\boldsymbol{E}^\dagger\boldsymbol{E} = I_4$, which leads to

$$\tilde{\boldsymbol{E}}^{\mathsf{T}}\tilde{\boldsymbol{E}} = \Re(\boldsymbol{E})^{\mathsf{T}}\Re(\boldsymbol{E}) + \Im(\boldsymbol{E})^{\mathsf{T}}\Im(\boldsymbol{E}) = I_4.$$

Thus, the last term in Eq. (26) is calculated to

$$\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B} + \boldsymbol{T}_1^{\mathsf{T}}\boldsymbol{T}_1 + \boldsymbol{T}_2^{\mathsf{T}}\boldsymbol{T}_2 + \boldsymbol{T}_3^{\mathsf{T}}\boldsymbol{T}_3 + \boldsymbol{T}_4^{\mathsf{T}}\boldsymbol{T}_4 = 2I_3.$$

Finally, the Schur complement (see Appendix B) is used to transform Eq. (26) to

$$\begin{bmatrix} (kI_{2n^2}-\boldsymbol{P})^{-1} & & & \tilde{\boldsymbol{E}}\boldsymbol{B} \\ & \begin{bmatrix} kI_{2n^2} & -\boldsymbol{P}\\ -\boldsymbol{P} & kI_{2n^2}\end{bmatrix}^{-1} & & \begin{bmatrix}\tilde{\boldsymbol{E}}\boldsymbol{T}_1\\ \tilde{\boldsymbol{E}}\boldsymbol{T}_2\end{bmatrix} \\ & & \begin{bmatrix} kI_{2n^2} & \boldsymbol{P}\\ \boldsymbol{P} & kI_{2n^2}\end{bmatrix}^{-1} & \begin{bmatrix}\tilde{\boldsymbol{E}}\boldsymbol{T}_3\\ \tilde{\boldsymbol{E}}\boldsymbol{T}_4\end{bmatrix} \\ \boldsymbol{B}^{\top}\tilde{\boldsymbol{E}}^{\top} & [\boldsymbol{T}_1^{\top}\tilde{\boldsymbol{E}}^{\top}\ \boldsymbol{T}_2^{\top}\tilde{\boldsymbol{E}}^{\top}] & [\boldsymbol{T}_3^{\top}\tilde{\boldsymbol{E}}^{\top}\ \boldsymbol{T}_4^{\top}\tilde{\boldsymbol{E}}^{\top}] & \tau\boldsymbol{S}+(2k+\epsilon-1)I_3 \end{bmatrix} \geq 0, \tag{29}$$

which is obviously an LMI with respect to the variables $\boldsymbol{E}$, ϵ, and τ. As a result, the original problem is equivalently written by

$$\begin{aligned} \min_{\mathbf{E},\epsilon,\tau} \quad & \epsilon, \\ \text{s.t.} \quad & (\mathbf{E},\epsilon,\tau) \in \mathcal{N}_1, \end{aligned} \tag{30}$$

where $\mathcal{N}_1$ is the following non-convex set:

$$\begin{aligned} \mathcal{N}_1 := \{\ (\boldsymbol{E},\epsilon,\tau) \mid \Phi(\boldsymbol{E}) \geq 0,\ \langle\!\langle I_n|\boldsymbol{E} = \langle\!\langle I_2|, \\ \text{LMI (29)},\ 0 \leq \epsilon \leq 1,\ \text{rank}\Phi(\boldsymbol{E}) = 1\ \}. \end{aligned} \tag{31}$$

B. The second stage: rank-minimization

Let us consider a closely related problem

$$\begin{aligned} \min_{\mathbf{E},\epsilon,\tau} \quad & \text{rank}\Phi(\boldsymbol{E}) + \gamma\epsilon, \\ \text{s.t.} \quad & (\mathbf{E},\epsilon,\tau) \in \mathcal{N}, \end{aligned} \tag{32}$$

where $\mathcal{N}$ is a convex set given by

$$\begin{aligned} \mathcal{N} := \{\ (\boldsymbol{E},\epsilon,\tau) \mid \Phi(\boldsymbol{E}) \geq 0,\ \langle\!\langle I_n|\boldsymbol{E} = \langle\!\langle I_2|, \\ \text{LMI (29)},\ 0 \leq \epsilon \leq 1\ \}. \end{aligned} \tag{33}$$

The tuning parameter $\gamma > 0$ gives the relative weight between the two objectives rank$\Phi(\boldsymbol{E})$ and ϵ. This change of the problem is motivated by the fact that we can now apply some known heuristic methods for rank minimization problems, one of which is discussed below.

The minimization of the rank of a matrix subject to convex constraints is a ubiquitous problem in diverse areas of engineering such as control theory, system identification, statistics, signal processing, and computational geometry [26]. The general rank-minimization problem

$$\min\ \text{rank}\,X \quad \text{s.t. } X \in \mathcal{M} \text{ and } X \geq 0,$$

where $X \geq 0$ is the optimization matrix variable and $\mathcal{M}$ is a convex set denoting the constraints, is computationally NP-hard, thus we need to rely on heuristics. The log-det heuristic introduced and discussed in [25, 26, 27] provides an attractive approach. The heuristic is described as follows: The function $\log\det(X+\delta I)$ is used as a *smooth surrogate* for rank$X$ to yield

$$\min\ \log\det(X+\delta I) \quad \text{s.t. } X \in \mathcal{M} \text{ and } X \geq 0,$$

where $\delta > 0$ is a small regularization constant, and can be chosen to be on the order of the eigenvalues we can consider as zero. Although the surrogate function $\log\det(X+\delta I)$ is not convex, it is smooth on the positive definite cone and can be minimized locally using any local minimization method; we here use iterative linearization. Let X_i denote the i-th iterate of the optimization variable X. The linearization of $\log\det(X+\delta I)$ around X_i is given by

$$\begin{aligned} \log\det(X+\delta I) = \log\det(X_i+\delta I) \\ + \text{Tr}\left[(X_i+\delta I)^{-1}(X-X_i)\right], \end{aligned} \tag{34}$$

where we have used the fact that $\nabla\log\det X = X^{-1}$ when $X > 0$. Hence, we can minimize $\log\det(X+\delta I)$ over the constraint set $\mathcal{M}$ by iteratively minimizing the local linearization (34). This leads to

$$X_{i+1} = \underset{X\in\mathcal{M}}{\arg\min}\ \text{Tr}\left[(X_i+\delta I)^{-1}X\right].$$

The new optimal point is X_{i+1}. Since the log-det function is concave in X, at each iteration its value decreases, and the sequence of the function values generated converges to a local minimum of $\log\det(X+\delta I)$. This implies that the global optimal solution X_{opt} can be obtained by appropriately choosing an initial point X_0 (see Fig. 1).

The above procedure is directly applicable to the case where the objective function is replaced by rank$X + \gamma\epsilon$ with $\epsilon \in [0,1]$ an additional variable and $\gamma > 0$ a constant. Therefore, the rank-minimization problem (32) is replaced by

$$\begin{aligned} \min_{\mathbf{E},\epsilon,\tau} \quad & \log\det(\Phi(\boldsymbol{E})+\delta I_{2n}) + \gamma\epsilon, \\ \text{s.t.} \quad & (\mathbf{E},\epsilon,\tau) \in \mathcal{N}. \end{aligned} \tag{35}$$

The local or global optimal solution of this problem is obtained by solving the following iterative SDP:

$$\begin{aligned} & (\boldsymbol{E}_{i+1},\epsilon_{i+1},\tau_{i+1}) \\ & \quad = \underset{(\mathbf{E},\epsilon,\tau)\in\mathcal{N}}{\text{argmin}} \Big\{ \text{Tr}\left[(\Phi(\boldsymbol{E}_i)+\delta I_{2n})^{-1}\Phi(\boldsymbol{E})\right] + \gamma\epsilon \Big\}. \end{aligned} \tag{36}$$

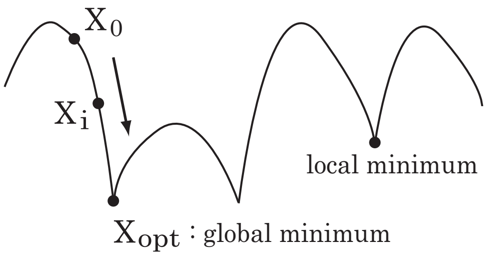

FIG. 1: The log-det function and a convergence of the iteration variable X_i.

Note that the convergence point of this algorithm is very sensitive to an initial point $\boldsymbol{E}_0$. (We do not need to specify ϵ_0 and τ_0, since ϵ_i and τ_i are not used to calculate $(\boldsymbol{E}_{i+1}, \epsilon_{i+1}, \tau_{i+1})$.)

We now provide an important theorem that connects the replaced problem (35) with the original problem (30).

Theorem 4. If the local (global) optimal solution $(\boldsymbol{E}_{\rm opt}, \epsilon_{\rm opt}, \tau_{\rm opt})$ of the problem (35) satisfies $\mathrm{rank}\Phi(\boldsymbol{E}_{\rm opt}) = 1$, then it coincides with the local (global) optimal solution of the problem (30).

Proof. The global optimal solution of (35) satisfies

$$\begin{aligned}\log\det(\Phi(\boldsymbol{E}_{\rm opt}) + \delta I_{2n}) + \gamma\epsilon_{\rm opt} \\ \leq \log\det(\Phi(\boldsymbol{E}) + \delta I_{2n}) + \gamma\epsilon, \end{aligned} \tag{37}$$

for all $(\boldsymbol{E}, \epsilon, \tau) \in \mathcal{N}$. From the assumption and Corollary 3, the nonzero eigenvalue of $\Phi(\boldsymbol{E}_{\rm opt})$ is $\dim\mathbb{C}^r = 2$, which yields $\det(\Phi(\boldsymbol{E}_{\rm opt}) + \delta I_{2n}) = (2+\delta)\delta^{2n-1}$. Also, any $\boldsymbol{E} \in \mathcal{X}_1(\mathbb{C}^2, \mathbb{C}^n)$ satisfies $\det(\Phi(\boldsymbol{E}) + \delta I_{2n}) = (2+\delta)\delta^{2n-1}$. As a result, Eq. (37) reduces to

$$\epsilon_{\rm opt} \leq \epsilon, \quad \forall (\boldsymbol{E}, \epsilon, \tau) \in \mathcal{N}_1,$$

where $\mathcal{N}_1$ is defined in Eq. (31). This implies that $(\boldsymbol{E}_{\rm opt}, \epsilon_{\rm opt}, \tau_{\rm opt})$ is indeed the global optimal solution of (30). We can prove the same fact for any local optimal solution by considering local regions of $\mathcal{N}$ and $\mathcal{N}_1$. ■

Clearly, the above theorem can be extended to the general case of r. Therefore, the optimal (suboptimal) solution of the original problem (20) is obtained by equivalently transforming (relaxing) it to a problem of the form (30) and solving a related rank-minimization problem via the same heuristic.

Finally, let us discuss choosing an initial point $\boldsymbol{E}_0$ of the algorithm (36) such that the iteration variable $(\boldsymbol{E}_i, \epsilon_i, \tau_i)$ converges to a local optimal solution $(\boldsymbol{E}_N, \epsilon_N, \tau_N)$ with $\mathrm{rank}\Phi(\boldsymbol{E}_N) = 1$. We here make the following observation; an initial point $\boldsymbol{E}_0$ that also satisfies $\mathrm{rank}\Phi(\boldsymbol{E}_0) = 1$ might be a good candidate for the above requirement to be satisfied. From the proof of Lemma 2, this implies

$$\boldsymbol{E}_0 = E_0 \otimes E_0^*, \quad E_0^\dagger E_0 = I_2. \tag{38}$$

Actually, in many practical cases, we observe that an initial point of the form (38) converges to a feasible local optimal solution. This fact will be seen in Section VI.

C. Exact optimal encoder for general qubit inputs

We here consider the general qubit input $|\phi\rangle = [\mathrm{e}^{\mathrm{i}\alpha}\cos\beta, \sin\beta] \in \mathbb{C}^2$ $(\alpha, \beta \in \mathbb{R})$ and outline the equivalent transformation of the constraint (21). We first note that the input vector $|\phi\rangle|\phi\rangle^*$ is represented in terms of a monomial vector as follows:

$$|\phi\rangle|\phi\rangle^* = \begin{bmatrix} \cos^2\beta \\ \mathrm{e}^{\mathrm{i}\alpha}\sin\beta\cos\beta \\ \mathrm{e}^{-\mathrm{i}\alpha}\sin\beta\cos\beta \\ \sin^2\beta \end{bmatrix} = \boldsymbol{U} \begin{bmatrix} x_1^2 + x_2^2 \\ \sqrt{2}x_1x_3 \\ \sqrt{2}x_2x_3 \\ x_3^2 \end{bmatrix} =: \boldsymbol{U}|x\rangle\!\rangle_U,$$

where the real variables $x_1, x_2, x_3 \in \mathbb{R}$ are defined as

$$x_1 = \cos\beta\cos\alpha, \;\; x_2 = \cos\beta\sin\alpha, \;\; x_3 = \sin\beta,$$

and $\boldsymbol{U}$ is a unitary matrix given by

$$\boldsymbol{U} := \begin{bmatrix} 1 & & & \\ & 1/\sqrt{2} & \mathrm{i}/\sqrt{2} & \\ & 1/\sqrt{2} & -\mathrm{i}/\sqrt{2} & \\ & & & 1 \end{bmatrix}.$$

Then, similar to the previous case, defining a $2n^2 \times 4$ real matrix

$$\tilde{\boldsymbol{E}}' := \begin{bmatrix} \Re(\boldsymbol{EU}) \\ \Im(\boldsymbol{EU}) \end{bmatrix},$$

the output purity is expressed by

$$P(\mathcal{A}, \mathcal{E}, |\phi\rangle) = {}_U\langle\!\langle x|\tilde{\boldsymbol{E}}'^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}'|x\rangle\!\rangle_U.$$

Consequently, the original max-min problem is equal to the minimization of $\epsilon \in [0, 1]$ subject to the conditions $\boldsymbol{E} \in \mathcal{X}_1(\mathbb{C}^2, \mathbb{C}^n)$ and

$$\begin{aligned} p'(x) := {}_U\langle\!\langle x|\Big[\tilde{\boldsymbol{E}}'^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}' + (\epsilon-1)I_4\Big]|x\rangle\!\rangle_U \geq 0, \\ \forall x_1, x_2, x_3 \in \mathbb{R}. \end{aligned}$$

Since $p'(x)$ is a fourth order homogeneous polynomial with respect to the three variables (x_1, x_2, x_3), the Hilbert's lemma (iii) in Eq. (A1) can be applied; the nonnegativity of $p'(x)$ is equivalent to the condition

$$p'(x) \text{ is an SOS with respect to } (x_1, x_2, x_3).$$

Then, as the SOS decomposition of $p'(x)$ implies the existence of a positive semidefinite matrix $\boldsymbol{Q}' \geq 0$ satisfying $p'(x) = {}_U\langle\!\langle x|\boldsymbol{Q}'|x\rangle\!\rangle_U$, the matrix $\tilde{\boldsymbol{E}}'^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}' + (\epsilon-1)I_4$ is related to $\boldsymbol{Q}'$ by

$$\begin{aligned} \boldsymbol{Q}' = \sum_i \boldsymbol{T}_i'^{\mathsf{T}}\Big[\tilde{\boldsymbol{E}}'^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}} + (\epsilon-1)I_4\Big]\boldsymbol{T}_i'' \\ + \sum_i \boldsymbol{T}_i''^{\mathsf{T}}\Big[\tilde{\boldsymbol{E}}'^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}} + (\epsilon-1)I_4\Big]\boldsymbol{T}_i' \\ + \sum_i \tau_i'\boldsymbol{S}_i' \geq 0, \end{aligned}$$

with certain matrices $\boldsymbol{T}'_i, \boldsymbol{T}''_i, \boldsymbol{S}'_i$, and additional scalar variables $\tau'_i \in \mathbb{R}$. The above nonlinear matrix inequality with respect to the variables $\boldsymbol{E}, \epsilon$, and τ'_i is further transformed to an LMI using the same technique shown in Section IV-A. As before, we then consider the problem of minimizing $\mathrm{rank}\Phi(\boldsymbol{E}) + \gamma\epsilon$ subject to the LMI obtained above and the linear constraints $\Phi(\boldsymbol{E}) \geq 0$, $\langle\!\langle I_n | \boldsymbol{E} = \langle\!\langle I_2 |$, and $0 \leq \epsilon \leq 1$. If the optimal solution of this problem satisfies $\mathrm{rank}\Phi(\boldsymbol{E}_{\rm opt}) = 1$, then it is also the optimal solution of the original problem (20) with $r = 2$.

V. SUBOPTIMAL SOLUTION IN HIGHER DIMENSIONAL CODE SPACE

In the general case $r \geq 3$, nonnegativity of a homogeneous polynomial no longer implies the existence of its SOS decomposition (this remarkable equivalence holds only in the cases (A1)). However, the SOS characterization can still be used as a sufficient condition; that is, the first constraint in Eq. (20) is *relaxed* to

$$p''(x) := \langle\!\langle x | \big[\tilde{\boldsymbol{E}}''^{\mathsf{T}} \boldsymbol{P} \tilde{\boldsymbol{E}}'' + (\epsilon - 1) I_{r^2} \big] | x \rangle\!\rangle \quad \text{is an SOS with respect to } (x_1, \ldots, x_{2r-1}), \tag{39}$$

where $\tilde{\boldsymbol{E}}'' \in \mathbb{R}^{2n^2 \times r^2}$ is an appropriately defined real matrix variable that is linear to $\boldsymbol{E}$, and $|x\rangle\!\rangle \in \mathbb{R}^{r^2}$ is an appropriately defined real monomial vector of $x_1, \ldots, x_{2r-1}$. The SOS condition (39) equivalently leads to an LMI as seen before, and consequently, we have a problem of the form (30) that can be tackled via the log-det heuristic. Note again that Eq. (39) is only a sufficient condition for the inequality $p''(x) \geq 0$ to be satisfied for all $(x_1, \ldots, x_{2r-1})$. Thus, any feasible solution $(\tilde{\boldsymbol{E}}'', \epsilon)$ satisfying Eq. (39) is included in the original set of solutions. Therefore, the suboptimal error computed from the relaxed problem, $\epsilon_{\rm sub}$, is always bigger than or equal to the exact optimal error $\epsilon_{\rm opt}$. This indicates that the suboptimal output purity, $P_{\rm sub}(\mathcal{A}) = 1 - \epsilon_{\rm sub}$, gives a lower bound of the optimal purity:

$$P(\mathcal{A}) = 1 - \epsilon_{\rm opt} \geq 1 - \epsilon_{\rm sub} = P_{\rm sub}(\mathcal{A}).$$

An important fact to be noticed is that, as pointed out in [24], the gap between the set of nonnegative polynomials and the set of polynomials with an SOS decomposition is considered to be small in a practical situation. Hence, we expect that $P_{\rm sub}(\mathcal{A})$ is a good approximation to $P(\mathcal{A})$.

VI. EXAMPLES

A. The bit-flip channel

The quantum bit-flip channel with flipping probability p is given by

$$\mathcal{S}(\mathbb{C}^2) \ni \rho \to \mathcal{T}_1 \rho = p \sigma_x \rho \sigma_x + q\rho \in \mathcal{S}(\mathbb{C}^2),$$

where $p + q = 1$ and $\sigma_x = |0\rangle\langle 1| + |1\rangle\langle 0|$. We here consider the double bit-flip channel $\mathcal{A}_{\rm bf} = \mathcal{T}_1^{\otimes 2}$:

$$\mathcal{S}(\mathbb{C}^4) \ni \rho \to \rho' = \mathcal{A}_{\rm bf}\rho = \sum_{i=1}^{4} A_i \rho A_i^\dagger \in \mathcal{S}(\mathbb{C}^4),$$
$$A_1 = p\, \sigma_x \otimes \sigma_x,\ A_2 = \sqrt{pq}\, \sigma_x \otimes I_2,$$
$$A_3 = \sqrt{pq}\, I_2 \otimes \sigma_x,\ A_4 = q\, I_2 \otimes I_2.$$

The matrix form of the double bit-flip channel, $\boldsymbol{A}_{\rm bf} = \sum_i A_i \otimes A_i^* \in \mathcal{X}(\mathbb{C}^4, \mathbb{C}^4)$, is represented by

$$\boldsymbol{A}_{\rm bf} = \left[\begin{array}{cc|cc} qA_4 & \sqrt{pq}A_3 & \sqrt{pq}A_2 & pA_1 \\ \sqrt{pq}A_3 & qA_4 & pA_1 & \sqrt{pq}A_2 \\ \hline \sqrt{pq}A_2 & pA_1 & qA_4 & \sqrt{pq}A_3 \\ pA_1 & \sqrt{pq}A_2 & \sqrt{pq}A_3 & qA_4 \end{array} \right].$$

In particular, we set $p = 0.1$; then, for example, $k = 2$ satisfies the condition (27): $kI_{32} - \boldsymbol{P} > 0$.

We here assume that the input is a real-valued qubit: $|\phi\rangle \in \mathbb{R}^2$. Then, an exact local or global optimal encoder $\boldsymbol{E}_{\rm opt} \in \mathcal{X}_1(\mathbb{R}^2, \mathbb{C}^4)$ is computed by the algorithm (36) under the condition $\mathrm{rank}\Phi(\boldsymbol{E}_{\rm opt}) = 1$. A strong convergence property of $\boldsymbol{E}_i$ is observed when the SDP parameters are set to $\delta = 0.01$ and $\gamma = 15$. We usually need 90 iterations of the SDP; hence we denote the convergence point by $(\boldsymbol{E}_{90}, \epsilon_{90}, \tau_{90})$. Note again that $\boldsymbol{E}_{90}$ must be of the form $\boldsymbol{E}_{90} = E_{90} \otimes E_{90}^*$ due to the rank condition $\mathrm{rank}\Phi(\boldsymbol{E}_{90}) = 1$. Regarding the initial point $\boldsymbol{E}_0$, we follow the idea mentioned in the last paragraph of Section IV-B and examine some initial points of the form (38) to find the global optimal solution.

First, we randomly choose two initial points as $\boldsymbol{E}_0^{(j)} = E_0^{(j)} \otimes E_0^{(j)*}$ $(j = 1, 2)$, where the Kraus operators $E_0^{(1)}$ and $E_0^{(2)}$ are given by

$$E_0^{(1)} = \frac{1}{\sqrt{10}} \begin{bmatrix} 2 & 0 \\ \sqrt{2} & -\sqrt{6} \\ \sqrt{3} & 1 \\ 1 & \sqrt{3} \end{bmatrix},\quad E_0^{(2)} = \frac{1}{\sqrt{10}} \begin{bmatrix} \sqrt{2} & 0 \\ \sqrt{3} & -\sqrt{6} \\ 1 & \sqrt{2} \\ 2 & \sqrt{2} \end{bmatrix},$$

respectively. Then, the corresponding convergence points are respectively given by $\boldsymbol{E}_{90}^{(j)} = E_{90}^{(j)} \otimes E_{90}^{(j)*}$ $(j = 1, 2)$, where

$$\{\ E_{90}^{(1)},\ E_{90}^{(2)}\ \} = \left\{ \begin{bmatrix} 0.5308 & -0.4672 \\ 0.5308 & -0.4672 \\ 0.4672 & 0.5308 \\ 0.4672 & 0.5308 \end{bmatrix},\ \begin{bmatrix} 0.5274 & -0.4710 \\ 0.5274 & -0.4710 \\ 0.4710 & 0.5274 \\ 0.4710 & 0.5274 \end{bmatrix} \right\}.$$

In both cases, the convergence value of the error is given by $\epsilon_{90} = 0.18$. In view of the structure of $E_{90}^{(1)}$ and $E_{90}^{(2)}$, we expect that the encoder $\mathcal{E}^{(\rm a)}$:

$$\boldsymbol{E}^{(\rm a)} = E^{(\rm a)} \otimes E^{(\rm a)*},\ E^{(\rm a)} = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \tag{40}$$

will be a local optimal solution and provide a local minimum of the error, $\epsilon^{(\mathrm{a})} = 0.18$, for all $\alpha \in [0, 2\pi)$. In fact, for the input $|\phi\rangle = [x_1,\ x_2]^{\mathsf{T}} = [\cos\varphi,\ \sin\varphi]^{\mathsf{T}}$ and the encoder $\mathcal{E}^{(\mathrm{a})}$, the output purity (18) is reduced to

$$P(\mathcal{A}_{\mathrm{bf}}, \mathcal{E}^{(\mathrm{a})}, |\phi\rangle) = 1 - 2pq\big[\cos(2\varphi + 2\alpha)\big]^2,$$

which takes the minimum value

$$P^{(\mathrm{a})}_{\min} = \min_{\varphi} P(\mathcal{A}_{\mathrm{bf}}, \mathcal{E}^{(\mathrm{a})}, |\phi\rangle) = 1 - 2pq = 0.82.$$

Hence, as expected above, the local minimum of the error is $\epsilon^{(\mathrm{a})} = 1 - 0.82 = 0.18$. This result clarifies that the optimal encoder depends on the worst-case input as $\alpha = -\varphi_{\mathrm{worst}} + n\pi/2$, where n is any integer. As a summary, the encoder

$$\begin{aligned}\mathcal{E}^{(\mathrm{a})}:\ \ & |\phi\rangle = [x_1,\ x_2]^{\mathsf{T}} \\ & \to |\phi_c\rangle = E^{(\mathrm{a})}|\phi\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} x_1\cos\alpha - x_2\sin\alpha \\ x_1\cos\alpha - x_2\sin\alpha \\ x_1\sin\alpha + x_2\cos\alpha \\ x_1\sin\alpha + x_2\cos\alpha \end{bmatrix}\end{aligned}$$

is locally optimal for all $\alpha \in [0, 2\pi)$.

We next try following two initial points: $\boldsymbol{E}^{(j)}_0 = E^{(j)}_0 \otimes E^{(j)*}_0$ $(j = 3, 4)$, where

$$E^{(3)}_0 = \frac{1}{\sqrt{10}}\begin{bmatrix} 2 & 0 \\ \sqrt{2} & \sqrt{6} \\ \sqrt{3} & -1 \\ -1 & \sqrt{3} \end{bmatrix},\quad E^{(4)}_0 = \frac{1}{\sqrt{10}}\begin{bmatrix} \sqrt{3} & -1 \\ \sqrt{2} & \sqrt{6} \\ 2 & 0 \\ 1 & -\sqrt{3} \end{bmatrix}.$$

Then, the corresponding convergence points are respectively given by $\boldsymbol{E}^{(j)}_{90} = E^{(j)}_{90} \otimes E^{(j)*}_{90}$ $(j = 3, 4)$ where

$$\begin{aligned}&\{\ E^{(3)}_{90},\ E^{(4)}_{90}\ \} \\ &= \left\{ \begin{bmatrix} 0.6935 & -0.1377 \\ 0.1377 & 0.6935 \\ 0.6935 & -0.1377 \\ 0.1377 & 0.6935 \end{bmatrix},\ \begin{bmatrix} 0.4636 & -0.5341 \\ 0.5341 & 0.4636 \\ 0.5341 & 0.4636 \\ 0.4636 & -0.5341 \end{bmatrix} \right\}.\end{aligned}$$

Although they have a similar structure, there is a large gap between the corresponding convergence values of the error ϵ:

$$\epsilon^{(3)}_{90} = 0.18,\quad \epsilon^{(4)}_{90} = 0.2952.$$

The structure of $E^{(3)}_{90}$ and $E^{(4)}_{90}$ suggests that the encoders $\boldsymbol{E}^{(\mu)} = E^{(\mu)} \otimes E^{(\mu)*}$ $(\mu = \mathrm{b}, \mathrm{c})$ with

$$\begin{aligned}&\{\ E^{(\mathrm{b})},\ E^{(\mathrm{c})}\ \} \\ &= \left\{ \frac{1}{\sqrt{2}}\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \\ \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix},\ \frac{1}{\sqrt{2}}\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \\ \sin\alpha & \cos\alpha \\ \cos\alpha & -\sin\alpha \end{bmatrix} \right\}\end{aligned}$$

are locally optimal for all $\alpha \in [0, 2\pi)$. Actually, the output purity (18) with the above encoders and the input $|\phi\rangle = [\cos\varphi,\ \sin\varphi]^{\mathsf{T}}$ are respectively calculated as

$$\begin{aligned}P(\mathcal{A}_{\mathrm{bf}}, \mathcal{E}^{(\mathrm{b})}, |\phi\rangle) &= 1 - 2pq[\cos(2\varphi + 2\alpha)]^2, \\ P(\mathcal{A}_{\mathrm{bf}}, \mathcal{E}^{(\mathrm{c})}, |\phi\rangle) &= 1 - 4pq(p^2 + q^2)[\cos(2\varphi + 2\alpha)]^2.\end{aligned}$$

Thus, their minimum values are

$$\begin{aligned}P^{(\mathrm{b})}_{\min} &= 1 - 2pq = 0.82, \\ P^{(\mathrm{c})}_{\min} &= 1 - 4pq(p^2 + q^2) = 0.7048,\end{aligned}$$

irrespective of α. The minimums are attained when $\cos(2\varphi + 2\alpha) = \pm 1$, as in the case of $\mathcal{E}^{(\mathrm{a})}$. We also see the following inequality:

$$P^{(\mathrm{b})}_{\min} - P^{(\mathrm{c})}_{\min} = 2pq(1 - 2p)^2 \geq 0.$$

Therefore, the encoders $\mathcal{E}^{(\mathrm{a})}$ and $\mathcal{E}^{(\mathrm{b})}$ achieve the same local minimum of the error, whereas $\mathcal{E}^{(\mathrm{c})}$ is inferior to those channels for all p.

Combining the entire set of investigations presented above with other numerical results that were omitted for brevity, we maintain that $\epsilon_{\mathrm{opt}} = 0.18$ is the global minimum and that the optimal purity is thus given by

$$P(\mathcal{A}_{\mathrm{bf}}) = 1 - \epsilon_{\mathrm{opt}} = 0.82.$$

The solutions $\mathcal{E}^{(\mathrm{a})}$ and $\mathcal{E}^{(\mathrm{b})}$ are typical optimal encoders that yield the above optimal purity.

Remark 1. In the Kraus representation, the output state is given by $\rho' = \sum_i A_i E|\phi\rangle\langle\phi|E^\dagger A_i^\dagger$. Intuitively, in order for the purity of ρ' to have a large value, the encoder E should be chosen so that the vectors $\{A_i E|\phi\rangle\}$ are close to each other. Actually, if all of them are parallel, the output state is pure. In this sense, $\mathcal{E}^{(\mathrm{a})}$ is a physically reasonable encoder because the vectors $A_1 E^{(\mathrm{a})}|\phi\rangle$ and $A_3 E^{(\mathrm{a})}|\phi\rangle$ are parallel to $A_2 E^{(\mathrm{a})}|\phi\rangle$ and $A_4 E^{(\mathrm{a})}|\phi\rangle$, respectively. The encoders $E^{(\mathrm{b})}$ and $E^{(\mathrm{c})}$ also satisfy such relations. In contrast, if we choose $E = |00\rangle\langle 0| + |11\rangle\langle 1|$ in Eq. (3), the four vectors $A_i E|\phi\rangle$ $(i = 1, \ldots, 4)$ differ from each other and span a linear space of dimension 4. This is indeed a bad encoder since the minimum output purity in this case is calculated as $p[\rho'] = (p^2 + q^2)^2 \approx 0.67$, which is clearly less than the optimal purity $P(\mathcal{A}_{\mathrm{bf}}) = 0.82$.

Remark 2. We again maintain that $\boldsymbol{E}_0$ satisfying $\mathrm{rank}\Phi(\boldsymbol{E}_0) = 1$ is a good initial point. Actually, within our investigation, we have observed that such an initial point always converges to a rank-one solution by appropriately choosing the SDP parameters δ and γ. However, for initial points with the rank more than one, it is easy to find a bad example of $\boldsymbol{E}_0$ such that $\mathrm{rank}\Phi(\boldsymbol{E}_{90}) = 1$ is not achieved for any δ and γ. For instance, if we choose $\Phi(\boldsymbol{E}_0) = (1/4)I_8$, then $\boldsymbol{E}_i$ always converges to a solution satisfying $\mathrm{rank}\Phi(\boldsymbol{E}_{90}) = 2$. Another reason of the emphasis is based on the following observation. Once we obtain a rank-one solution using an initial point with the rank more than one, then we always have found a rank-one initial point that converges to the same solution. In other words, it is considered that any rank-one solution is available by choosing a rank-one initial point appropriately. For example, $\boldsymbol{E}_i$ starting from $\boldsymbol{E}_0 = 0.5\boldsymbol{E}^{(1)}_0 + 0.3\boldsymbol{E}^{(2)}_0 + 0.2\boldsymbol{E}^{(3)}_0$ converges into a rank-one solution of the form (40).

B. The amplitude damping channel

The amplitude damping channel describes the dissipation of a quantum state into equilibrium due to coupling with its environment. The Kraus representation of the channel is given by

$$\mathcal{S}(\mathbb{C}^2) \ni \rho \to \mathcal{T}_2\rho = H_1\rho H_1^\dagger + H_2\rho H_2^\dagger \in \mathcal{S}(\mathbb{C}^2),$$

where

$$H_1 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{p} \end{bmatrix},\ H_2 = \begin{bmatrix} 0 & \sqrt{1-p} \\ 0 & 0 \end{bmatrix}.$$

The parameter $p \in (0,1)$ represents the rate of dissipation. We consider the double amplitude damping channel $\mathcal{A}_{\rm ad} = \mathcal{T}_2^{\otimes 2}$:

$$\mathcal{S}(\mathbb{C}^4) \ni \rho \to \rho' = \mathcal{A}_{\rm ad}\rho = \sum_{i=1}^4 A_i\rho A_i^\dagger \in \mathcal{S}(\mathbb{C}^4),$$
$$A_1 = H_1 \otimes H_1,\ A_2 = H_1 \otimes H_2,$$
$$A_3 = H_2 \otimes H_1,\ A_4 = H_2 \otimes H_2.$$

The matrix form of the channel, $\boldsymbol{A}_{\rm ad} = \sum_i A_i \otimes A_i^* \in \mathcal{X}(\mathbb{C}^4, \mathbb{C}^4)$, is given by

$$\boldsymbol{A}_{\rm ad} = \left[\begin{array}{cc|cc} A_1 & \sqrt{1-p}A_2 & \sqrt{1-p}A_3 & (1-p)A_4 \\ O_4 & \sqrt{p}A_1 & O_4 & \sqrt{p(1-p)}A_3 \\ \hline O_4 & O_4 & \sqrt{p}A_1 & \sqrt{p(1-p)}A_2 \\ O_4 & O_4 & O_4 & pA_1 \end{array}\right],$$

where O_4 denotes the 4×4 zero matrix. In particular, we consider the case of $p = 0.1$ and set $k = 4$, which leads to $kI_{32} - \boldsymbol{P} > 0$.

Our goal is to obtain the optimal encoder under the condition $|\phi\rangle \in \mathbb{R}^2$, in which case $\boldsymbol{E}_{\rm opt} \in \mathcal{X}_1(\mathbb{R}^2, \mathbb{C}^4)$. The iteration variable $\boldsymbol{E}_i$ of the algorithm (36) is initialized to $\boldsymbol{E}_0$ of the form (38), and the SDP parameters are set to $\delta = 0.01$ and $\gamma = 6.1$. In order to find a rank-one convergence point, we usually need 500 iterations of the SDP; we thus denote the convergence point by $(\boldsymbol{E}_{500}, \epsilon_{500}, \tau_{500})$.

First, let us take the initial points $\boldsymbol{E}_0^{(1)}$ and $\boldsymbol{E}_0^{(3)}$, which have appeared in the bit-flip channel case. Then, the corresponding convergence points are respectively given by $\boldsymbol{E}_{500}^{(j)} = E_{500}^{(j)} \otimes E_{500}^{(j)*}$ $(j = 5, 6)$ with the Kraus operators

$$\{\ E_{500}^{(5)},\ E_{500}^{(6)}\ \}$$
$$= \left\{ \begin{bmatrix} 0.6555 & 0.2503 \\ 0.4174 & -0.9045 \\ 0.5741 & 0.2822 \\ 0.2583 & 0.1989 \end{bmatrix},\ \begin{bmatrix} 0.5803 & -0.2273 \\ 0.4668 & 0.8745 \\ 0.5568 & -0.2843 \\ 0.3679 & -0.3209 \end{bmatrix} \right\}.$$

In both cases, the convergence value of the error is given by $\epsilon_{500} = 0.18$. Unlike the case of bit-flip channel, the above solutions do not have a simple structure of the matrix entries, which is highly important for a physical realization of encoding process. To obtain a simple solution, let us carry out the algorithm with an initial point that has a specific matrix form itself. As a typical example, we consider the following initial point:

$$\boldsymbol{E}_0^{(\rm d)} = E_0^{(\rm d)} \otimes E_0^{(\rm d)*},\ \ E_0^{(\rm d)} = \begin{bmatrix} \cos\alpha & 0 \\ 0 & \cos\beta \\ \sin\alpha & 0 \\ 0 & \sin\beta \end{bmatrix}.$$

Then, for any $\alpha \in [0, 2\pi)$ and $\beta \in (0, \pi/2)$, $\boldsymbol{E}_i$ converges to

$$\boldsymbol{E}_{500}^{(\rm d)} = E_{500}^{(\rm d)} \otimes E_{500}^{(\rm d)*},\ \ E_{500}^{(\rm d)} = \begin{bmatrix} \cos\alpha & 0 \\ 0 & 1 \\ \sin\alpha & 0 \\ 0 & 0 \end{bmatrix},$$

with $\epsilon_{500}^{(\rm d)} = 0.18$. This encoder is locally optimal for all $\alpha \in [0, 2\pi)$. Actually, the output purity (18) for the encoder-error process $\mathcal{A}_{\rm ad}\mathcal{E}_{500}^{(\rm d)}$ with the input $|\phi\rangle = [x_1,\ x_2]^\mathsf{T}$ is calculated as

$$P(\mathcal{A}_{\rm ad}, \mathcal{E}_{500}^{(\rm d)}, |\phi\rangle) = 1 - 2p(1-p)(x_1^2\sin^2\alpha + x_2^2)^2, \quad (41)$$

and thus, its minimum is $P_{\rm min}^{(\rm d)} = 1 - 2p(1-p) = 0.82$ when $|\phi\rangle = [0, 1]^\mathsf{T}$ irrespective of α.

We also observe the following similar convergence:

$$E_0^{(\rm e)} = \begin{bmatrix} \cos\alpha & 0 \\ \sin\alpha & 0 \\ 0 & \cos\beta \\ 0 & \sin\beta \end{bmatrix} \ \to\ E_{500}^{(\rm e)} = \begin{bmatrix} \cos\alpha & 0 \\ \sin\alpha & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

with $\epsilon_{500}^{(\rm e)} = 0.18$ for all $\alpha \in [0, 2\pi)$ and $\beta \in (0, \pi/2)$. The output purity $P(\mathcal{A}_{\rm ad}, \mathcal{E}_{500}^{(\rm e)}, |\phi\rangle)$ has the same form as Eq. (41), thus the encoder $\mathcal{E}_{500}^{(\rm e)}$ is also locally optimal for all $\alpha \in [0, 2\pi)$.

Finally, let us choose an initial point of the form

$$\boldsymbol{E}_0^{(\rm f)} = E_0^{(\rm f)} \otimes E_0^{(\rm f)*},\ \ E_0^{(\rm f)} = \begin{bmatrix} \cos\alpha & 0 \\ 0 & \cos\beta \\ 0 & \sin\beta \\ \sin\alpha & 0 \end{bmatrix}. \quad (42)$$

We then observe a somewhat complicated convergence depending on (α, β) as follows. When α takes a small number, e.g., $\alpha = 0.2$ (any β can be taken), the algorithm does not cause a variation in $\boldsymbol{E}_i$, and only ϵ_i changes into 0.18. That is, we obtain the local optimal solution

$$\boldsymbol{E}_{500}^{(\rm f_1)} = E_{500}^{(\rm f_1)} \otimes E_{500}^{(\rm f_1)*},\ \ E_{500}^{(\rm f_1)} = \begin{bmatrix} \cos\alpha & 0 \\ 0 & \cos\beta \\ 0 & \sin\beta \\ \sin\alpha & 0 \end{bmatrix}.$$

On the other hand, when $\alpha \approx \pi/2$, another type of convergence occurs. For example when choosing $\alpha = 1.3$, $\boldsymbol{E}_i$ converges to

$$\boldsymbol{E}_{500}^{(\rm f_2)} = E_{500}^{(\rm f_2)} \otimes E_{500}^{(\rm f_2)*},\ \ E_{500}^{(\rm f_2)} = \begin{bmatrix} 0.6893 & 0 \\ 0 & \cos\beta \\ 0 & \sin\beta \\ 0.7245 & 0 \end{bmatrix},$$

with $\epsilon_{500}^{(\mathrm{f}_2)} = 0.18$. To further understand this complex structure of the solution, we provide an analytical investigation of the output purity $P(\mathcal{A}_{\mathrm{ad}}, \mathcal{E}_0^{(\mathrm{f})}, |\phi\rangle)$ in Appendix C. However, we reemphasize that a lucid advantage of our method to search an optimal solution is that it does not require any analytic examination on the max-min optimization problem of the output purity, which is in general extremely hard.

Based on the above investigations, we maintain that $\epsilon_{\mathrm{opt}} = 0.18$ is the global minimum and that $\mathcal{E}_{500}^{(\mu)}$ ($\mu = \mathrm{d}, \mathrm{e}, \mathrm{f}_1, \mathrm{f}_2$) are the optimal encoders. Therefore, the optimal purity is given by

$$P(\mathcal{A}_{\mathrm{ad}}) = 1 - \epsilon_{\mathrm{opt}} = 0.82.$$

VII. CONCLUSION

In this paper, we presented a tractable computational algorithm for designing a quantum encoder that maximizes the worst-case output purity of a given decohering channel over all possible pure inputs. We cast the problem as a max-min optimization problem (minimization over all pure inputs, and maximization over all pure state preserving encoders). Although this problem is computationally very hard to solve due to the non-convexity property, our algorithm computes the exact optimal solution for codespace of dimension two. Moreover, we showed an extended version of the above algorithm that computes a lower bound of the optimal purity for the general class of codespaces.

We believe that the proposed computational approach provides a powerful method that is also applicable to other problems in quantum encoding and fault-tolerant quantum information transmissions. For example, following the same techniques presented in this paper, we can prove that a quantum error correction problem with the minimum fidelity criterion considered in [5, 18] is transformed or relaxed to a convex optimization problem systematically; we are then able to obtain the optimal or suboptimal solution using SDP. This result will be reported soon.

Acknowledgments

We wish to thank P. Parrilo for pointing out the SOS characterization. NY would like to acknowledge stimulating discussions with S. Hara and H. Siahaan. MF thanks M. Yanagisawa for helpful discussions. This work was supported in part by the Grants-in-Aid for JSPS fellows No.06693.

APPENDIX A: PROOF OF EQ. (25)

Let us consider a real polynomial function $p(x)$ in n variables $x = (x_1, \ldots, x_n)$ of the form:

$$p(x) = \sum_k c_k x_1^{k_1} \cdots x_n^{k_n}, \quad c_k \in \mathbb{R},$$

where the sum is over n-tuples $k = (k_1, \ldots, k_n)$ satisfying $\sum_{i=1}^n k_i = m$. This function is called the *homogeneous polynomial* of degree m in n variables. A homogeneous polynomial satisfies $p(\lambda x_1, \ldots, \lambda x_n) = \lambda^m p(x_1, \ldots, x_n)$. We now state the famous Hilbert's theorem. Let $P_{n,m}$ be the set of nonnegative homogeneous polynomials of degree m in n variables. Let $\Sigma_{n,m}$ be the set of homogeneous polynomials $p(x)$ that has an SOS decomposition $p(x) = \sum_i h_i(x)^2$, where $h_i(x)$ are homogeneous polynomials of degree $m/2$. Then, $P_{n,m} = \Sigma_{n,m}$ holds only in the following cases:

$$\text{(i) } n = 2 \quad \text{(ii) } m = 2 \quad \text{(iii) } n = 3,\ m = 4. \qquad \text{(A1)}$$

For more detailed description on this problem, see [31].

Now, Eq. (24) has the following form:

$$p(x) = [x_1^2\ \sqrt{2}x_1x_2\ x_2^2]\boldsymbol{H}\begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix} \geq 0, \quad \forall x_1, x_2 \in \mathbb{R}, \qquad \text{(A2)}$$

where $\boldsymbol{H} = (h_{ij})$ is a real 3×3 symmetric matrix. The function $p(x)$ is a homogeneous polynomial with respect to two variables x_1 and x_2 (and degree $m = 4$). Therefore, from the Hilbert's formula (i) in Eq. (A1), the constraint (A2) is equivalent to the condition

$$p(x) \text{ is an SOS with respect to } x_1 \text{ and } x_2.$$

Moreover, it can be shown that the existence of an SOS decomposition is equivalent to the existence of a positive semidefinite matrix $\boldsymbol{Q} = (q_{ij}) \geq 0$ such that

$$p(x) = z(x)^{\mathsf{T}}\boldsymbol{Q}z(x), \qquad \text{(A3)}$$

where $z(x)$ is a vector of monomials of degree equal to $\deg(p)/2 = 2$. Comparing Eq. (A3) with (A2), we set $z(x) = [x_1^2,\ \sqrt{2}x_1x_2,\ x_2^2]^{\mathsf{T}}$. Then, the equality $z(x)^{\mathsf{T}}\boldsymbol{H}z(x) = z(x)^{\mathsf{T}}\boldsymbol{Q}z(x)$ yields

$$h_{11} = q_{11},\ h_{12} = q_{12},\ \ h_{13} + h_{22} = q_{13} + q_{22},$$
$$h_{23} = q_{23},\ h_{33} = q_{33},$$

which leads to

$$\boldsymbol{Q} = \begin{bmatrix} h_{11} & h_{12} & q_{13} \\ h_{12} & h_{22} + h_{13} - q_{13} & h_{23} \\ q_{13} & h_{23} & h_{33} \end{bmatrix} \geq 0.$$

As a result, Eq. (A2) is equivalent to the following matrix inequality:

$$\begin{bmatrix} h_{11} & h_{12} & 0 \\ h_{12} & h_{22} + h_{13} & h_{23} \\ 0 & h_{23} & h_{33} \end{bmatrix} + \tau \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \geq 0,$$

where $\tau := q_{13} \in \mathbb{R}$ is an additional optimization variable. The above inequality can be expressed as

$$\boldsymbol{H} + \boldsymbol{S}_1^{\mathsf{T}}\boldsymbol{H}\boldsymbol{S}_2 + \boldsymbol{S}_2^{\mathsf{T}}\boldsymbol{H}\boldsymbol{S}_1 - \boldsymbol{S}_3^{\mathsf{T}}\boldsymbol{H}\boldsymbol{S}_4 - \boldsymbol{S}_4^{\mathsf{T}}\boldsymbol{H}\boldsymbol{S}_3 + \tau\boldsymbol{S} \geq 0,$$

where

$$\boldsymbol{S}_1 = \begin{bmatrix} 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ \boldsymbol{S}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 \end{bmatrix},$$

$$\boldsymbol{S}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ \boldsymbol{S}_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ \boldsymbol{S} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

From the above discussion, the constraint (24) is equivalently transformed to

$$\begin{aligned}
&\boldsymbol{B}^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{B} + (\epsilon-1)I_3 + \tau\boldsymbol{S} \\
&\quad + \boldsymbol{S}_1^{\mathsf{T}}\Big[\boldsymbol{B}^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{B} + (\epsilon-1)I_3\Big]\boldsymbol{S}_2 \\
&\quad + \boldsymbol{S}_2^{\mathsf{T}}\Big[\boldsymbol{B}^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{B} + (\epsilon-1)I_3\Big]\boldsymbol{S}_1 \\
&\quad - \boldsymbol{S}_3^{\mathsf{T}}\Big[\boldsymbol{B}^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{B} + (\epsilon-1)I_3\Big]\boldsymbol{S}_4 \\
&\quad - \boldsymbol{S}_4^{\mathsf{T}}\Big[\boldsymbol{B}^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{B} + (\epsilon-1)I_3\Big]\boldsymbol{S}_3 \geq 0.
\end{aligned}$$

As $\boldsymbol{S}_1^{\mathsf{T}}\boldsymbol{S}_2 = O$ and $\boldsymbol{S}_3^{\mathsf{T}}\boldsymbol{S}_4 = O$, we obtain Eq. (25):

$$\begin{aligned}
&\boldsymbol{B}^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{B} + (\epsilon-1)I_3 + \tau\boldsymbol{S} \\
&+ \boldsymbol{T}_1^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{T}_2 + \boldsymbol{T}_2^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{T}_1 \\
&- \boldsymbol{T}_3^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{T}_4 - \boldsymbol{T}_4^{\mathsf{T}}\tilde{\boldsymbol{E}}^{\mathsf{T}}\boldsymbol{P}\tilde{\boldsymbol{E}}\boldsymbol{T}_3 \geq 0,
\end{aligned}$$

where the matrices $\boldsymbol{T}_i$ $(i = 1, \ldots, 4)$ are defined as

$$\boldsymbol{T}_1 := \boldsymbol{B}\boldsymbol{S}_1, \ \boldsymbol{T}_2 := \boldsymbol{B}\boldsymbol{S}_2, \ \boldsymbol{T}_3 := \boldsymbol{B}\boldsymbol{S}_3, \ \boldsymbol{T}_4 := \boldsymbol{B}\boldsymbol{S}_4.$$

APPENDIX B: THE SCHUR COMPLEMENT

The Schur complement is a powerful tool that transforms a convex but nonlinear constraint with respect to matrix variables into an equivalent LMI. Its derivation is very easy; assuming $A > 0$, we have a matrix equation of the form:

$$\begin{bmatrix} I & O \\ -B^{\dagger}A^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ B^{\dagger} & C \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ O & I \end{bmatrix} = \begin{bmatrix} A & O \\ O & C - B^{\dagger}A^{-1}B \end{bmatrix}.$$

Hence, the following relation holds:

$$\begin{bmatrix} A & B \\ B^{\dagger} & C \end{bmatrix} \geq 0 \ \Leftrightarrow \ \begin{cases} A > 0 \\ C - B^{\dagger}A^{-1}B \geq 0. \end{cases}$$

This is termed the Schur complement. In order to see the usefulness, let us consider a nonlinear constraint of a matrix variable X: $I - X^{\dagger}X \geq 0$. The Schur complement states that the constraint is equivalent to

$$\begin{bmatrix} I & X \\ X^{\dagger} & I \end{bmatrix} \geq 0,$$

which is obviously an LMI.

APPENDIX C: AN ANALYTIC INVESTIGATION OF THE PURITY-OPTIMIZATION PROBLEM

We here give an observation on the purity-optimization problem where the error channel is $\mathcal{A}_{\rm ad}$ and the encoder is $\mathcal{E}_0^{({\rm f})}\rho = E_0^{({\rm f})}\rho E_0^{({\rm f})*}$ with $E_0^{({\rm f})}$ given in Eq. (42). The output purity $P(\mathcal{A}_{\rm ad}, \mathcal{E}_0^{({\rm f})}, |\phi\rangle) = {\rm Tr}\,[\mathcal{A}_{\rm ad}\mathcal{E}_0^{({\rm f})}(|\phi\rangle\langle\phi|)^2]$ with the input $|\phi\rangle = [x_1, \ x_2]^{\mathsf{T}} \in \mathbb{R}^2$ is then calculated to

$$\begin{aligned}
P(\alpha, \beta, x_1) = 1 - 2pq\Big[&(1 + \sin 2\alpha \sin 2\beta - 2pq\sin^4\alpha)x_1^4 \\
&- (1 + \cos 2\alpha + \sin 2\alpha \sin 2\beta)x_1^2 + 1\Big].
\end{aligned}$$

First, let us consider the case where α takes a small number. Especially when $\alpha = 0$, $P(0, \beta, x_1) = 1 - 2pq(x_1^2 - 1)^2$ is a concave function with respect to x_1. Thus, the minimum is given by $P_{\rm min} = P(0, \beta, 0) = 1 - 2pq$ at $x_1 = 0$. This fact is still true for $\alpha \approx 0$; the function $P(\alpha, \beta, x_1)$ is concave and takes the minimum $1 - 2pq$ at $x_1 = 0$ without respect to the values of α and β. This is the reason why α and β do not have specific optimal values and the iterative SDP initialized with $\alpha \approx 0$ does not renew these parameters. On the other hand, when $\alpha = \pi/2$, the output purity becomes

$$P(\pi/2, \beta, x_1) = 1 - 2pq\Big[(p^2 + q^2)x_1^4 + 1\Big],$$

which obviously takes the minimum at $x_1 = 1$. Moreover, for $\alpha \approx \pi/2$ the function $P(\alpha, \beta, x_1)$ is still concave and takes the minimum $P(\alpha, \beta, 1) = 1 + 4p^2q^2\sin^2\alpha(\sin^2\alpha - 1/pq)$. Unlike the case of $\alpha \approx 0$, this function must be further maximized with respect to α. For this reason, there is a specific optimal value of α, whereas β does not affect the optimality.

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge (2000).

[2] K. Kraus, *States, Effects and Operations, Fundamental Notions of Quantum Theory* (Academic, Berlin, 1983).

[3] P. Shor, Phys. Rev. A 52, 2493 (1995).

[4] A. M. Steane, Phys. Rev. Lett. 77, 793 (1996).

[5] E. Knill and R. Laflamme, Phys. Rev. A 55, 900 (1997).

[6] D. A. Lidar, I. L. Chuang, and K. B. Whaley, Phys. Rev. Lett. 81, 2594 (1998).

[7] D. A. Lidar, D. Bacon, and K. B. Whaley, Phys. Rev. Lett. 82, 4556 (1999).

[8] A. Shabani and D. A. Lidar, Phys. Rev. A 72, 042303 (2005).

[9] P. Zanardi and D. A. Lidar, Phys. Rev. A 70, 012315 (2004).

[10] L. Vandenberghe and S. Boyd, SIAM Review 38, 49 (1996).

[11] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in systems and control theory*, (SIAM, Philadelphia, 1994).

[12] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, Phys. Rev. Lett. 88, 187904 (2002).

[13] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, Phys. Rev. A 69, 022308 (2004).

[14] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, Phys. Rev. A 71, 032333 (2005).

[15] J. Eisert, P. Hyllus, O. Guhne, and M. Curty, Phys. Rev. A 70, 062317 (2004).

[16] R. O. Vianna and A. C. Doherty, Phys. Rev. A 74, 052306 (2006).

[17] H. M. Wiseman and A. C. Doherty, Phys. Rev. Lett. 94, 070405 (2005).

[18] N. Yamamoto, S. Hara and, K. Tsumura, Phys. Rev. A 71, 022322 (2005).

[19] A. S. Fletcher, P. W. Shor, and M. Z. Win, Phys. Rev. A 75, 012338 (2007).

[20] R. L. Kosut and D. A. Lidar, e-print quant-ph/0606078 (2006).

[21] A. Jamiolkowski, Rep. Math. Phys. 3, 275 (1972).

[22] P. Parrilo, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*, (Ph.D thesis, California Institute of Technology, Pasadena, CA, 2000).

[23] P. Parrilo, Mathematical Programming Ser. B, 96, 2 293/320 (2003).

[24] S. Prajna, A. Papachristodoulou, P. Seiler, and P. Parrilo, *Positive Polynomials in Control*, 273/292, (Springer, 2005).

[25] M. Fazel, H. Hindi, and S. P. Boyd, Proceedings of American Control Conference, June 2003.

[26] M. Fazel, *Matrix rank minimization with applications*, (Ph.D thesis, Stanford University, Stanford, CA, 2002).

[27] M. Fazel, H. Hindi, and S. P. Boyd, Proceedings of American Control Conference, June 2001.

[28] M. D. Choi, Linear Algebr. Appl. 10, 285 (1975).

[29] A. Fujiwara and P. Algoet, Phys. Rev. A 59, 3290 (1999).

[30] G. M. D'Ariano and P. Lo Presti, Phys. Rev. A 64, 042308 (2001).

[31] B. Reznick, *Contemporary Mathematics*, 253, 251/272, (AMS, 2000).

[32] In the absence of $\boldsymbol{B}$, we will need extra real scalar variables $\tau_1, \cdots, \tau_5$ in addition to $\boldsymbol{E}$ and ϵ in order to obtain an equivalent LMI, whereas our LMI (29) requires only one additional variable $\tau \in \mathbb{R}$.